*Data Feature Extraction Methods Based on Deep Learning*

1st Xiaoli Guo

Northeastern Electric Power University

Jilin, Jilin, China

243589657@qq.com

4th Zichong Feng

China Logistics Co., Ltd.

Harbin branch

Harbin, Heilongjiang, China

583168375@qq.com

2nd Minhan Wei

Northeastern Electric Power University

Jilin, Jilin, China

1272504062@qq.com

5th Yuhan Sun

Jilin City Investment Promotion Centre

Jilin, Jilin, China

632667701@qq.com

3rd Hao Jiang

Northeast Electric Power University

Jilin, Jilin, China

[1013107126@qq.com](mailto:1013107126@qq.com)

\*Corresponding author

*Abstract*—To address the deficiencies of frequently missed detections of small targets and low accuracy under occlusion scenarios, this paper proposes a feature extraction network based on attention mechanisms and multi-scale atrous convolutions. Initially, a multi-scale attention mechanism, based on SENet (MSENet), is designed to optimize the integration and weighting of channel information, thereby enhancing the model's focusing capability. Subsequently, to address the shortcomings of the existing SPPF module, an improved MDC module is constructed to maintain the resolution of feature maps while enlarging the receptive field. Finally, leveraging the MSENet and MDC modules, an improved YOLOv8 network structure is proposed that effectively enhances detection accuracy in complex environments. Validation through a series of experiments on a public dataset shows an increase of 1.95% in Precision and 0.95% in mAP compared to the baseline network, significantly improving detection performance in scenarios involving small targets, complex environments, and blurry backgrounds.

Keywords—Deep Learning, Feature Extraction, Object Detection

# Introduction

With the rise of motorcycles and electric bikes in cities, traffic safety issues have intensified. Inadequate laws and some riders' failure to wear helmets result in frequent electric bike accidents, particularly under adverse weather conditions, where head injuries are a major cause of death. Although helmets don't prevent accidents, they reduce fatalities and injuries. Motorcycle riders generally have higher safety awareness, but electric bike riders lack management, leading to safety gaps. The advancement of deep learning, especially in object detection, offers new solutions. Manual inspections are inefficient, while deep learning-based helmet detection enhances traffic safety supervision. However, existing research often focuses on general helmet detection and faces challenges with occlusion and complex backgrounds, making it less suitable for urban environments.

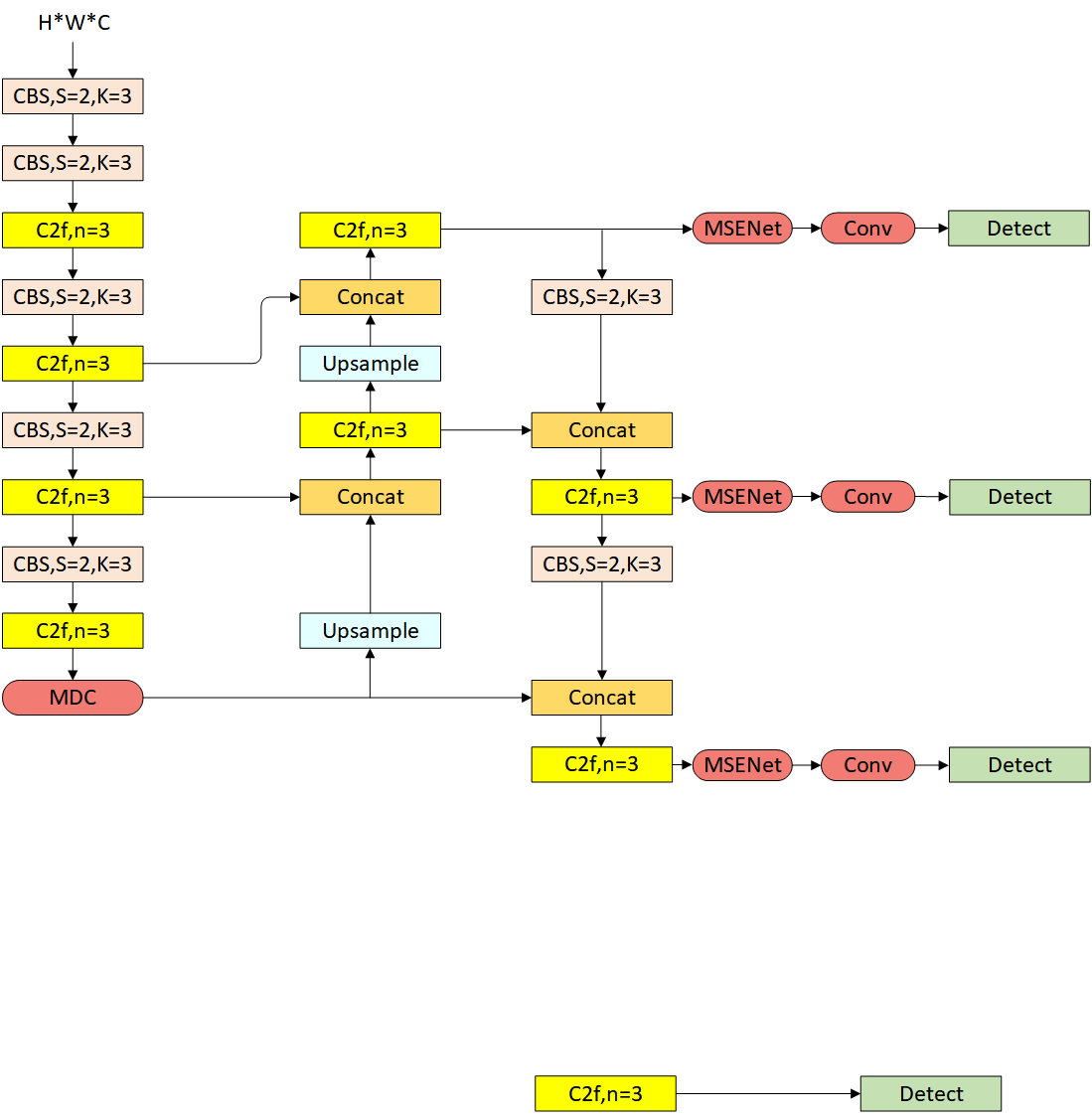
Regarding the detection of riding helmets, Wen et al.[1] proposed a Hough transform-based circular detection method, which first manually defines the threshold to calculate the image edges, and then applies the Hough transform to search for helmets within the circular region.Chiu et al.[2] used a Canny edge detector for helmet detection.Chiverton et al.[3] used background subtraction for detecting moving motorcycles, computed the image statistics to approximate the head region of the rider, and subsequently extracted helmet features from the isolated region using gradient histogram descriptors and finally classified them by support vector machine classifier. Silva et al.[4] extracted features using circular Hough transform and oriented gradient histogram descriptors for detection of motorcycle rider's head, and then determined whether they were wearing helmets or not using multilayer perceptron classifiers. Waranusat et al.[5] proposed a system that extracts features from region attributes by KNN classifier to first classify them as motorcycles or other moving targets, and then detect whether a helmet is worn or not by segmenting and classifying the head of a rider on a motorcycle. Similarly, Dahiya et al.[6] used background subtraction and object segmentation with a KNN classifier to detect bicycle riders.

Previous research on helmet detection for traffic cyclists has been relatively limited, with most similar studies and applications focusing on safety helmets in environments such as construction sites, factories, and buildings. Domestic scholars[7-11] have mainly focused on improving models such as YOLOv4, Faster R-CNN and SSD to improve the accuracy and speed of helmet detection. These improvements include methods such as re-generating a priori frames, adding feature detection layers, incorporating self-attention mechanisms, and using lightweight networks. In recent years, some scholars in China have begun to conduct research on helmet-wearing for traffic riding. Ran Duisheng's team[12] developed an improved YOLOv2 target detection algorithm, whose core innovation is to use MobileNetV2 to replace the traditional YOLOv2 backbone network and integrate the Efficient Channel Attention (ECA) mechanism as a way to enhance the feature representation ability of the model and reduce its complexity. Ruichen Xue et al.[13] proposed a helmet-wearing detection algorithm based on improved YOLOv3. The method effectively enhances feature extraction and detection of helmet-wearing by fusing channel and spatial attention modules and introducing a densely connected network. Qirui Li[14] proposed a two-stage helmet detection method based on YOLOv4 and YOLOv4-tiny cascade networks, which improves the detection accuracy of small targets by secondary filtering the detection results through category recoding and cascade networks.

This paper proposes a multi-scale attention mechanism based on SENet (MSENet) to enhance object detection accuracy through optimized channel information and weight distribution. To address information loss from the SPPF module in YOLOv8 and minimize MSENet's impact on feature extraction, we introduce a multi-scale dilated convolution (MDC) module. This module maintains feature map resolution while expanding the receptive field, effectively capturing multi-scale targets and improving detection efficiency in diverse scenarios.

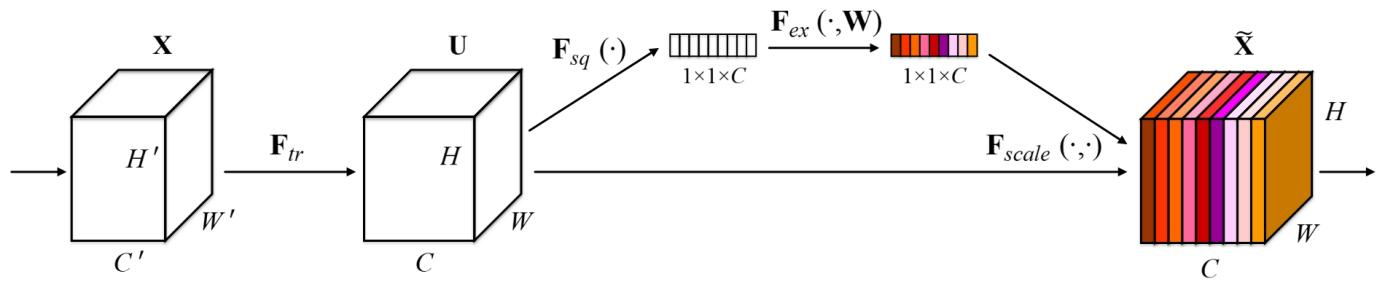
# Proposed Algorithm

Building upon the YOLOv8n framework, a feature extraction network incorporating multi-scale dilated convolutions and attention mechanisms was constructed, as illustrated in Figure 1.



1. Overall network structure
   1. *Improved multi-scale attention mechanism(MSENet)*

SENet(Squeeze-and-Excitation Network) represents an architectural refinement of convolutional neural networks (CNNs) through the integration of an attention mechanism. This design enhances the model's focus on critical information by adaptively emphasizing salient feature channels. Consequently, SENet significantly improves performance in tasks such as image processing, while concurrently alleviating computational load and mitigating overfitting issues. This approach underscores the potential of attention mechanisms to augment predictive accuracy and processing efficiency within neural architectures.



1. SENet module structure

As illustrated in Figure 2, the fundamental concept of SENet can be encapsulated by two key operations: Squeeze and Excitation:

The Squeeze operation involves global average pooling across the spatial dimensions of the feature map to generate a compressed feature vector that represents the global information of each channel, as shown in Equation:

(1)

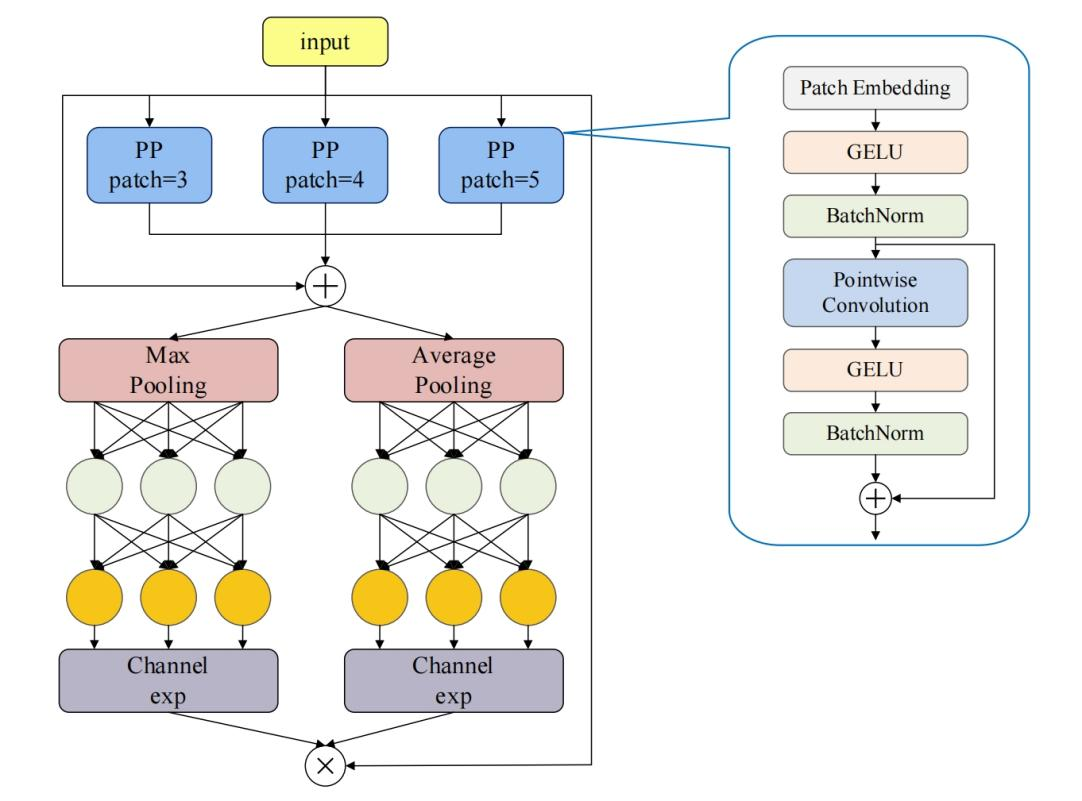
The Excitation operation leverages this compressed feature vector through a fully connected neural network, incorporating ReLU and sigmoid functions, to learn the importance weights of each feature channel, dynamically adjusting the feature responses, as illustrated in Equation:

(2)

The feature mapping operation multiplies the weights obtained from the Excitation operation with the original feature maps, thereby emphasizing important feature channels and suppressing less relevant ones.:

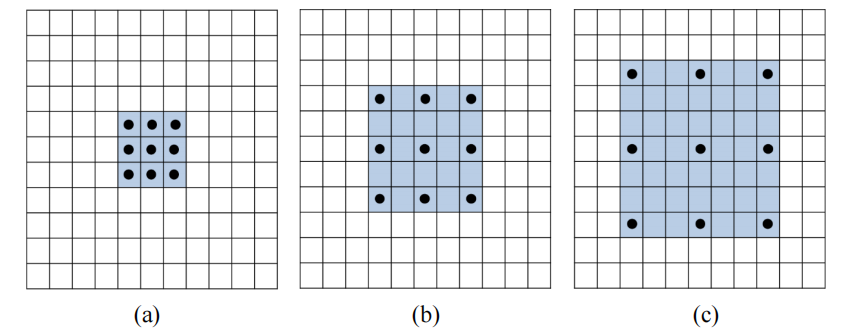
(3)

To address SENet's potential omission of key information when handling complex inter-channel correlations, an improved multi-scale attention mechanism (MSENet) is designed. The architecture consists of two parts: first, multi-scale pointwise convolutions address feature information loss and integrate inter-channel information. Second, features are compressed through two improved pooling layers, and channel weights are calculated via a fully connected layer. These weights are multiplied with the original feature map, forming an optimized attention mechanism that enhances detection performance, as shown in Figure 3.



1. Structure of MSENet module
   1. *Improved multi-scale dilated convolution module(MDC)*

Dilated convolution is a widely used technique in deep learning convolutional neural networks, particularly in image processing. By introducing gaps (dilation) within the traditional convolutional kernel, it allows the convolution operation to skip certain elements in the input data, thereby expanding the receptive field without increasing the number of parameters. This approach enhances the network's spatial coverage while avoiding the computational burden of large convolutional kernels. For instance, when the dilation size is 1, dilated convolution is equivalent to standard convolution. Dilated convolution effectively addresses the receptive field limitations in deep networks, reduces parameter count, decreases model complexity, and enhances the efficiency and predictive capability for large-scale datasets.



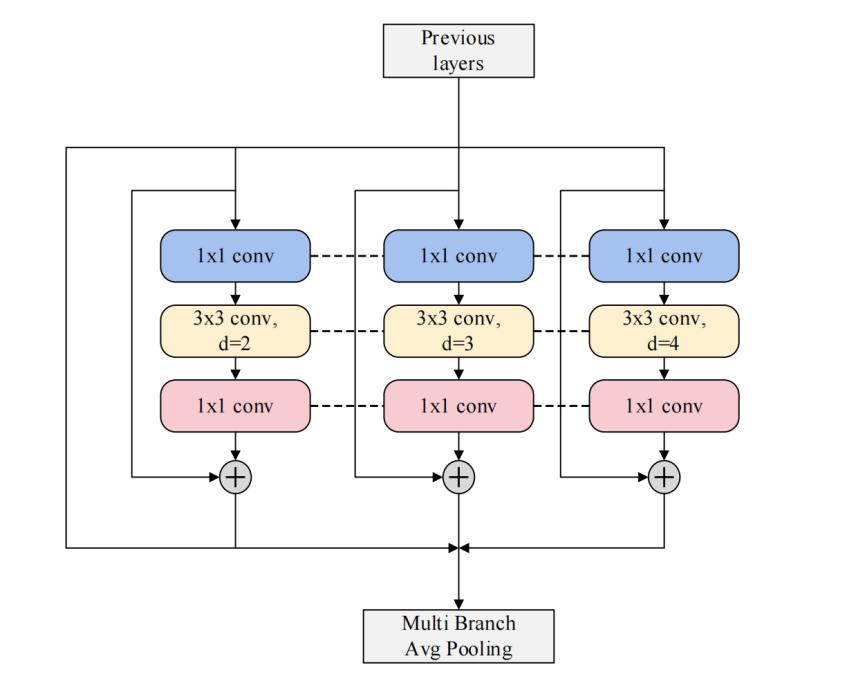
1. Convolution process of the null convolution structure

Figure 4 illustrates the dilated convolution process. The input image is represented by the outermost box, the convolutional kernel by black dots, and the receptive field after the convolution operation is shown in blue. In Figure 4(a), the dilation size is 1, resulting in a 3×3 receptive field; in Figure 4(b), the dilation size is 2, resulting in a 5×5 receptive field; and in Figure 4(c), the dilation size is 3, resulting in a 7×7 receptive field.

By adjusting the dilation size, the effective receptive field size of each element in the convolution kernel can be controlled. The calculation formula is as follows:

(4)

Here,𝑘𝑒 represents the receptive field, 𝑘 denotes the kernel size, and 𝑑 stands for the dilation rate. For example, in a traditional convolution operation with a kernel size of 3×3 (𝑘=3) and a dilation rate 𝑑=1, the receptive field is 3×3. When the dilation rate is 2, each kernel element skips one pixel in the input data, expanding the receptive field to 5×5. With a dilation rate of 3, the receptive field becomes 7×7. By adjusting the dilation rate, the receptive field can be expanded, enhancing the model's ability to capture multi-scale information.



1. Improved MDC module structure

The Multi-Scale Dilated Convolution (MDC) module is designed to expand the receptive field of the convolutional kernel while maintaining the size and integrity of the feature map. Unlike the pooling layers in the SPPF module, the MDC module uses dilated convolutions to more effectively extract and integrate information across channels, enhancing the ability to capture multi-scale information, as illustrated in Figure 5. Specifically, the MDC module employs dilations of different rates(d=2,3,4) to perform convolution operations, thus expanding the receptive field and capturing objects and contextual information at various scales. Additionally, by introducing residual connections, the original input is added to the output processed by 1x1 and dilated convolution layers, forming multi-scale feature maps. These are then concatenated and average pooled, enhancing the module's performance and expressive capability. Replacing the SPPF module with the MDC module and integrating it into the YOLOv8 backbone network allows the network to maintain image integrity while reducing the number of parameters.

# Simulation experiment and result analysis

## Experimental details

The experimental setup is configured as follows: the operating system is Ubuntu 18.04, and the CPU used is an Intel(R) Xeon(R) Gold 6240, equipped with a Tesla T4 GPU (16GB VRAM). The system has CUDA version 11.6 and CUDNN version 8.0.5 installed. The experiments utilized the OpenCV toolkit, with Tensorboard and Matplotlib for visualization. The proposed network was developed under the PyTorch framework, and the parameters were updated using the Stochastic Gradient Descent (SGD) method. The initial learning rate was set to 0.01, with a batch size of 8, and a total training period of 200 epochs. During the data processing stage, all image resolutions were uniformly adjusted to 640×640 pixels, and the entire network was trained end-to-end.

## Evaluation metrics

To validate the accuracy and detection speed of the proposed algorithm for helmet detection on riders, this study uses Precision and mean Average Precision (mAP) as evaluation metrics. The mAP, representing detection accuracy, is calculated based on Precision and Recall. Frames Per Second (FPS) is used to measure detection speed.

## Dataset

The dataset used in the experiment consists of publicly available online data sets, one of which is among the most widely used for helmet detection. Originally derived from the Myanmar video surveillance dataset, it encompasses a variety of standard traffic scenes captured in urban, suburban, and rural settings. The dataset includes 10,000 training images and 2,003 test images. The labels are categorized into three classes: "CrashHelmet," "Without," and "elecar."

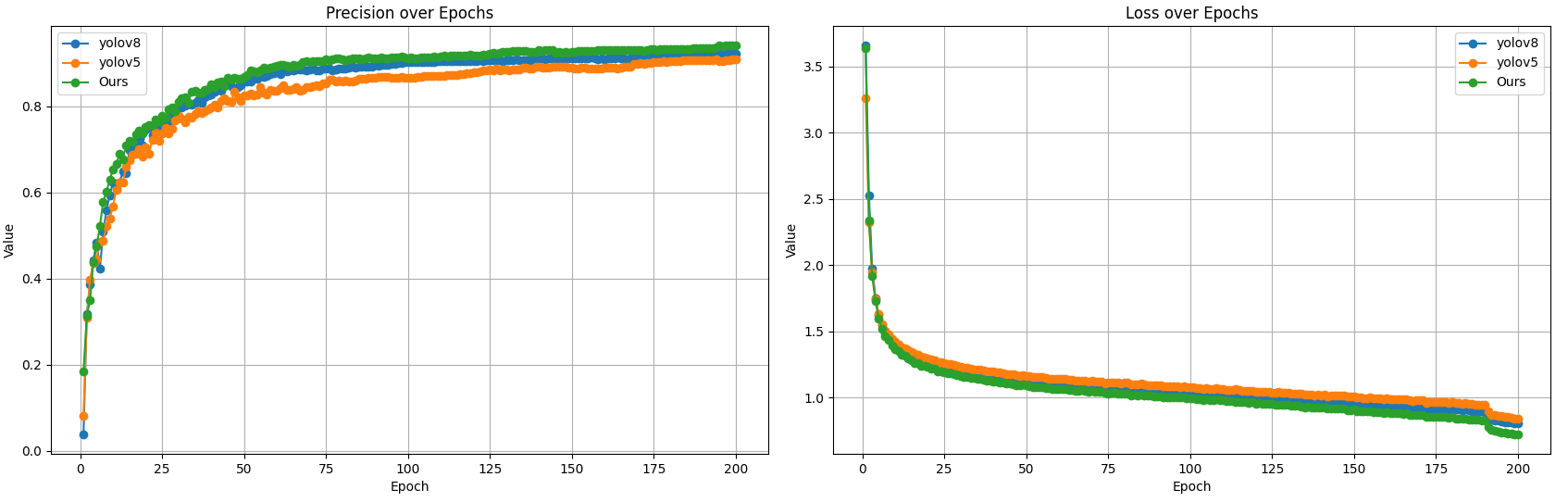
## Comparative experiments

This study selected representative and high-performance object detection algorithms to conduct experiments on a public helmet dataset, comparing them with the proposed YOLOv8n+MSNET+MDC (Ours) algorithm. Compared to the baseline YOLOv8n model, our approach improved accuracy by 1.95% and mAP by 0.95%, demonstrating superior detection performance. The experimental results are shown in Table I.

1. Comparison of experimental results with different Methods.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Precision** | **Recall** | **mAP@0.5** | **mAP@.5:.95** | **F1** |
| **YOLOv8n+MSNET+MDC**  **(Ours)** | **0.941** | **0.872** | **0.956** | **0.787** | **0.91** |
| YOLOv8-C2f-PPA | 0.927 | 0.857 | 0.947 | 0.766 | 0.89 |
| **YOLOv8n** | **0.923** | **0.86** | **0.947** | **0.762** | **0.89** |
| YOLOv8-RepNCSPELAN\_CAA | 0.915 | 0.834 | 0.933 | 0.741 | 0.87 |
| YOLOv8-CAA-HSFPN | 0.915 | 0.834 | 0.934 | 0.741 | 0.87 |
| YOLOv5n | 0.914 | 0.822 | 0.932 | 0.741 | 0.87 |
| YOLOv8-ELA-HSFPN | 0.91 | 0.826 | 0.93 | 0.74 | 0.87 |

From the Precision curve shown below, the YOLOv8n+MSNET+MDC improved model exhibits superior training performance. As the number of training iterations increases, the precision continuously improves. In the Loss curve, it is evident that the improved model has a lower loss value and converges faster than YOLOv8n. These results are illustrated in Figure 6.



1. Precision curve vs. Loss curve

## Ablation experiments

To further validate the effectiveness of each proposed module, ablation experiments were conducted. The accuracy comparison results are shown in Table 2. By comparing YOLOv8n+MSNET+MDC (Ours) with YOLOv8n, it is evident that the MSENet and MDC modules integrate well within the YOLOv8 network, achieving better performance than either the MSENet or MDC module alone.

1. Different modules in crash helmet detection.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **MSNET** | **MDC** | **Precision** | **Recall** | **mAP@0.5** | **mAP increase** | **Precision increase** |
| **YOLOv8n** | **－** | **－** | **0.923** | **0.86** | **0.95** | － | － |
| YOLOv8n+MSNET | ○ | － | 0.921 | 0.86 | 0.94 | 0.739% | 0.217% |
| YOLOv8n+MDC | － | ○ | 0.935 | 0.865 | 0.95 | 0.528% | 1.300% |
| **YOLOv8n+MSNET+MDC**  **(Ours)** | ○ | ○ | **0.941** | **0.872** | **0.96** | **0.950%** | **1.950%** |

After training the YOLOv8n+MSNET+MDC model, the trained model weights were used to detect helmet-wearing in the test set. Figure 7 shows the detection results of the YOLOv8n+MSNET+MDC algorithm in various scenarios. The left images display the original annotated detection boxes from the validation dataset, while the right images show the predicted detection boxes along with their corresponding confidence scores.



1. Detection effect of YOLOv8n+MSNET+MDC

# Conclusion

This paper proposes a deep learning-based method for data feature extraction. First, the MSENet multi-scale attention mechanism based on SENet optimizes channel information integration and weight allocation, enhancing the focus of the object detection model. Second, to address information loss in the SPPF module of the original YOLOv8, a multi-scale dilated convolution (MDC) strategy is introduced, expanding the receptive field to capture targets of various scales. Finally, by integrating the MSENet and MDC modules, an improved YOLOv8 network structure is proposed, enhancing helmet detection accuracy in complex environments.

##### Acknowledgements

The work of this paper is funded by: The Science and Technology Development Program of Jilin Province (No.20210203195SF).

##### References

1. Wen, C.Y., Chiu, S.H., Liaw, J.J., et al. "The safety helmet detection for ATM's surveillance system via the modified Hough transform." Proceedings of the IEEE 37th Annual 2003 International Carnahan Conference on Security Technology, 2003, IEEE, 2003: 364-369.
2. Chiu, C.C., Wang, C.Y., Ku, M.Y., et al. "Real time recognition and tracking system of multiple vehicles." Proceedings of the 2006 IEEE Intelligent Vehicles Symposium, IEEE, 2006: 478-483.
3. Chiverton, J. "Helmet presence classification with motorcycle detection and tracking." IET Intelligent Transport Systems, 2012, 6(3): 259-269.
4. Silva, R., Aires, K., Santos, T., et al. "Automatic detection of motorcyclists without helmet." Proceedings of the 2013 XXXIX Latin American Computing Conference (CLEI), IEEE, 2013: 1-7.
5. Waranusast, R., Bundon, N., Timtong, V., et al. "Machine vision techniques for motorcycle safety helmet detection." Proceedings of the 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), IEEE, 2013: 35-40.
6. Dahiya, K., Singh, D., Mohan, C.K. "Automatic detection of bike-riders without helmet using surveillance videos in real-time." Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, 2016: 3046-3051.
7. Zheng, X., Wang, S., Zhang, W., Zheng, J., Zhou, Y. "Safety helmet supervision system based on deep learning." Computer Systems Applications, 2021, 30(11): 118-126.
8. Jin, Y., Wu, X., Dong, H., Yu, L., Zhang, W. "Helmet-wearing detection algorithm based on improved YOLO v4." Computer Science, 2021, 48(11): 268-275.
9. Sun, G., Li, C., Zhang, H. "Helmet-wearing detection method integrating self-attention mechanism." Computer Engineering and Applications, 2022, 58(20): 5.
10. Li, M., Han, Q., Zhang, T., Wang, D. "Helmet detection method based on improved SSD." Computer Engineering and Applications, 2021, 57(08): 192-197.
11. Xu, D., Chen, Z. "Helmet-wearing detection method for electric bicycles based on improved YOLOv5s." Journal of Jinling Institute of Technology, 2022, 38(1): 7-14. DOI: 10.16515/j.cnki.32-1722/n.2022.01.002.
12. Ran, X., Chen, Z., Zhang, H. "Improved YOLOv2 algorithm for motorcycle helmet detection on roads." Electronic Measurement Technology, 2021, 44(24): 105-115.
13. Xue, R., Hao, Y., Zhang, Z., et al. "Helmet-wearing detection algorithm based on improved YOLOv3." Electronic Measurement Technology, 2021, 44(12): 115-120.
14. Li, Q. "Research on motorcycle helmet-wearing detection based on convolutional neural network." Master's thesis, Harbin University of Science and Technology, 2023.